Evaluation, CLIR Strategies and Metadata

Shibamouli Lahiri

December 6, 2010

Current Search Interface

Search Results

Dataset and Unit of Retrieval

- Original Cypriot documents (in Greek)
- English versions obtained using Google Translate

Each search result corresponds to a **page** of the original document.

- For example, **04_taifaco_AKYG6_231-655-0103** corresponds to the 103rd page of the document **04_taifaco_AKYG6_231-655**

Evaluation of English Search Results

- Independent judge
- Set of English queries from Wikipedia articles on Ancient Greece and Ancient Cyprus
- Topical instead of named-entity queries
- The search results give us an idea of the book contents

Archaic period, Battle of Corinth, Athenian Golden Age, Peloponnesian War, Classical Antiquity, Syro-Hittite, Assyrian, Phoenician, battle of Leuctra, Hellenistic period, Delian League, Mycenaean civilization, Lelantine War, Chalcis, helots, Battles of Salamis, Battles of Plataea, Epaminondas, Battle of Chaeronea, League of Macedon, Diadochi, Eudoxus of Cnidus, Knossos, Peace of Nicias, chiton, Hellas, Flaccus, Aulus Gellius

Query Set (Contd.)

Greek results but no English results

- poleis
- archon basileus
- Nea Nikomedia

Relevance Judgment

The goal is to create a gold-standard set of documents for the above queries. We use ternary relevance judgment:

- $2 = $ Relevant
- $1 = $ Marginally Relevant
- $0 = $ Not Relevant

Battle of Chaeronea	League of Macedon	Peace of Nicias
2	2	1
2	2	1
0	2	0
0	2	0
0	2	0
0	0	1
0	0	0
0	0	0
0	0	0
0	0	0

Evaluation

Metric	Value
P@1	0.7931
P@2	0.8276
P@3	0.8276
P@4	0.8276
P@5	0.8276
P@10	0.7744
MAP@10	0.8077
MRR	0.8276
NDCG@10	0.8297

- Language Model
- LDA-ESA (Roth and Klakow, 2010)

The Language Model approach requires **sentence alignment** and **translation probability estimation**. LDA-ESA takes Wikipedia-based **link information** into account.

- Alignment of Europarl and Perseus
- Probability Estimation (Brown, et al., 1993)

Document Structure

- **01_taifaco_proto:** Title of the Book, Publisher and Copyright, Preface, Introduction, Contents, Overall Description.

- **taifaco-arxaio001-154:** Testimonia concerning Cypriot authors and their fragments as found in other ancient authors in the original language (Ancient Greek or Latin). Each passage comes with various other readings found in manuscripts and papyri or suggested by scholars.

- **03_taifaco-arxaio155-229:** Continuation of taifaco-arxaio001-154.

- **05_taifaco_transl_005-229:** Modern Greek translation of different philosophers' works.

- **04_taifaco_AKYG6_231-655:** Modern Greek commentary on the works, a glossary, a list of abbreviations, bibliography, index, correlating table of fragments and an epilogue.
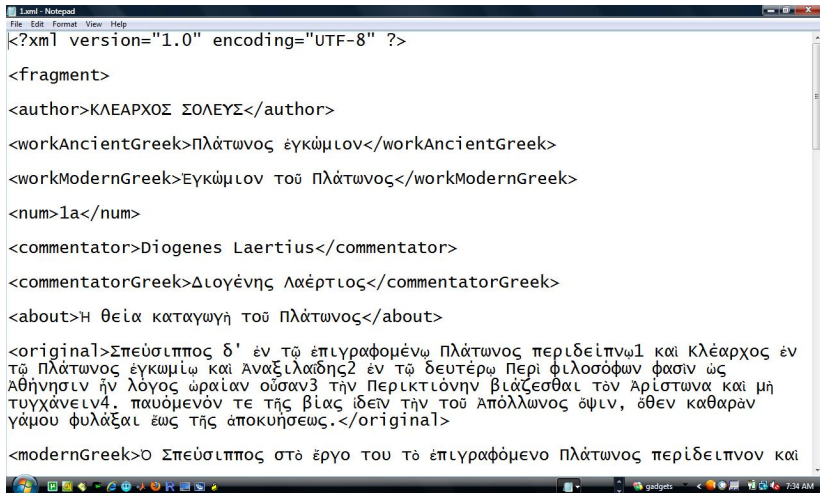
ΚΕΙΜΕΝΟ – ΜΕΤΑΦΡΑΣΗ
ΚΛΕΑΡΧΟΣ ΣΟΛΕΥΣ
Πλάτωνος ἐγκώμιον
F1a (2aW) Diogenes Laertius 3.2

Σπεύσιππος δ᾿ ἐν τῷ ἐπιγραφομένῳ Πλάτωνος περιδείπνῳ[1] καὶ Κλέαρχος ἐν τῷ Πλάτωνος ἐγκωμίῳ καὶ Ἀναξιλαΐδης[2] ἐν τῷ δευτέρῳ Περὶ φιλοσόφων φασὶν ὡς Ἀθήνησιν ἦν λόγος ὡραίαν οὖσαν[3] τὴν Περικτιόνην βιάζεσθαι τὸν Ἀρίστωνα καὶ μὴ τυγχάνειν[4]. παυόμενόν τε τῆς βίας ἰδεῖν τὴν τοῦ Ἀπόλλωνος ὄψιν, ὅθεν καθαρὰν γάμου φυλάξαι ἕως τῆς ἀποκυήσεως.

[1] περιδείπνω **B²P⁵**: περὶ δείπνω **F**: περὶ δείπνου **P¹QWH**   [2] Ἀναξιλαΐδης Cobet: ἀναξιλίδης **FP⁵ (Pᵖᶜ)**: ἀναξιλήδης **B**: ἀναξιλήδης **FP²**: ἀναξιάδης **P¹(Pᵃᶜ)W**: Ἀναξιλείδης Lang   [3] γάμου ante οὖσαν add. Marcovich   [4] ἐπὶ super τυγχάνειν add. **F²**

XML Formatted

```xml
<?xml version="1.0" encoding="UTF-8" ?>

<fragment>

<author>ΚΛΕΑΡΧΟΣ ΣΟΛΕΥΣ</author>

<workAncientGreek>Πλάτωνος ἐγκώμιον</workAncientGreek>

<workModernGreek>Ἐγκώμιον τοῦ Πλάτωνος</workModernGreek>

<num>1a</num>

<commentator>Diogenes Laertius</commentator>

<commentatorGreek>Διογένης Λαέρτιος</commentatorGreek>

<about>Η θεία καταγωγή τοῦ Πλάτωνος</about>

<original>Σπεύσιππος δ' ἐν τῷ ἐπιγραφομένῳ Πλάτωνος περιδείπνῳ1 καὶ Κλέαρχος ἐν
τῷ Πλάτωνος ἐγκωμίῳ καὶ Ἀναξιλαΐδης2 ἐν τῷ δευτέρῳ Περὶ φιλοσόφων φασὶν ὡς
Ἀθήνησιν ἦν λόγος ὡραίαν οὖσαν3 τὴν Περικτιόνην βιάζεσθαι τὸν Ἀρίστωνα καὶ μὴ
τυγχάνειν4. παυόμενόν τε τῆς βίας ἰδεῖν τὴν τοῦ Ἀπόλλωνος ὄψιν, ὅθεν καθαρὰν
γάμου φυλάξαι ἕως τῆς ἀποκυήσεως.</original>

<modernGreek>Ο Σπεύσιππος στὸ ἔργο του τὸ ἐπιγραφόμενο Πλάτωνος περίδειπνον καὶ
```
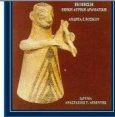
CyDAR Metadata Search

Check to select

☐	**Author:**	Select one ▾	Or type here...	AND ▾
☐	**Commentator:**	Select one ▾	Or type here...	AND ▾
☐	**Fragment:**	Select one ▾	Or type here...	AND ▾

Submit   Reset

- Metadata Search
- Sentence Alignment and Probability Estimation for the Language Model
- Comparison of Language Model with other approaches
- Implementation and evaluation of LDA-ESA

Acknowledgments

- Prof Zoe Stamatopoulou, Classics and Ancient Mediterranean Studies.
- Partha Mukherjee, IST.